CLAIMS

1. A method of using a computer system to identify a microbe inhabiting a host organism, comprising the steps of:

   a) obtaining sequence information from a plurality of sequences from at least one host organism; and

   b) searching a database of host organism genomic sequences to determine the presence or absence of said plurality of sequences in said database, wherein the absence of at least one of said sequences in said database indicates that said at least one sequence is a candidate sequence belonging to a microbe.

2. A method of using a computer system to identify a microbe inhabiting a host organism, comprising the steps of:

   a) obtaining sequence information from a library of genomic DNA from a host organism suspected of harboring a microbe;and

   b) searching a database of host organism genomic sequences from host organisms which do not harbor the microbe to determine the presence or absence of a sequence in said library in said database;
   wherein the absence of said sequence indicates that said sequence is a candidate microbe sequence.

3. A method of using a computer system to identify a microbe inhabiting a host organism, comprising the steps of:

   a) obtaining sequence information from a plurality of expressed sequences from at least one host organism; and

   b) searching a database of host organism genomic sequences to determine the presence or absence of said plurality of expressed sequences in said database, wherein the absence of at least one of said expressed sequences in said database indicates that said at least one sequence is a candidate sequence belonging to a microbe.

4. The method according to claims 1, 2, or 3, wherein said microbe is a symbiotic organism.

5. The method according to claim 4, wherein said microbe is a mutualistic organism, a commensal organism, or a parasitic organism.

6. The method according to claim 1, 2, or 3, wherein said microbe is a pathogenic organism.

7. The method according claims 1, 2, 3, wherein said plurality of sequences are compared to said database of host genomic sequences simultaneously.

8. A method of using a computer system to identify an intracellular pathogen, comprising the steps of:

a) obtaining sequence information from at least one host organism having a pathogenic condition;

b) identifying sequences from said at least one host organism which are not found in a plurality of host organisms not having said pathogenic condition;

c) comparing said sequences identified in step (b) with a plurality of sequences in a database of host genomic sequences; and

d) eliminating identified sequences which match said host genomic sequences, wherein any remaining sequences are identified as candidate pathogen sequences.

9. The method according to claim 8, wherein said identified sequences are compared simultaneously with sequences in said database of host genomic sequences.

10. The method according to claim 1 or 8, wherein said sequences are expressed sequences.

11. The method according to claim 1, 3, or 8, wherein said expressed sequences are EST sequences.

12. The method according to claim 1, 3, or 8, wherein said expressed sequences are cDNA sequences.

13. The method according to claim 1, 2, 3, or 8, wherein said host organism is an animal.

14. The method according to claim 13, wherein said animal is a mammal.

15. The method according to claim 14, wherein said mammal is a human.

16. The method according to claim 13, wherein said animal is an insect, bird, or a fish.

17. The method according to claim 1, 2, 3, or 8, wherein said host organism is a microrganism, a fungus, or a plant.

18. The method according to claim 11, wherein said candidate sequence is identified by comparing sequences in a database of expressed sequences with said sequences in said genomic database.

19. The method according to claim 8, wherein said expressed sequence is identified using a differential gene expression assay.

20. The method according to claim 19, wherein said differential gene expression assay is selected from the group consisting of SAGE, cDNA representational difference analysis, and suppression subtraction analysis.

21. The method according to claim 8, wherein said at least one sequence is identified using a subtractive hybridization method.

22. The method according to claim 21, wherein said subtractive hybridization method is representational difference analysis.

23. The method according to claim 1, 2, 3, or 8, wherein said candidate sequence is used as a query sequence to search a database of microbial sequences.

24. The method according to claim 23, wherein said microbial sequences include viral sequences.

25. The method according to claim 1, 2, 3, or 8, wherein any of: vector sequences, repetitive sequences, mitochondrial sequences, non-host species sequences, known host organism sequences, and combinations thereof are eliminated from the genomic database comprising sequences from the host organism.

26. The method according to claim 1, 2, 3, or 8, wherein said searching is performed iteratively using progressively smaller word sizes.

27. The method according to claim 1, 2, 3, or 8, wherein said candidate sequence is used to probe a library of sequences including sequences from at least one microbe.

28. The method according to claim 27, wherein a sequence identified by said probe is used to express a peptide.

29. The method according to claim 6 or 8, wherein said pathogen is an infectious disease organism.

30. The method according to claim 6 or 8, wherein said pathogen is associated with a pathogenic condition selected from the group consisting of an inflammatory disease, an autoimmune disease, and a cell proliferative disease.

31. The method according to claim 30, wherein said disease is selected from the group consisting of sarcoidosis, inflammatory bowel disease, atherosclerosis, multiple sclerosis, rheumatoid arthritis, type I diabetes mellitus, lupus erythematosus, Hodgkin's disease, and bronchioalveolar carcinoma.

32. The method according to claim 1, 2, 3, or 8, wherein said candidate sequence is used to produce a peptide.

33. The method according to claim 1, 2, 3, or 8 wherein said candidate sequence is operably linked to a promoter sequence in an expression vector.

34. The method according to claim 32, wherein said peptide is administered to the host organism in an amount effective to generate a protective immune response.

35. The method according to claim 33, wherein said expression vector is administered to the host organism in an amount effective to generate a protective immune response.

36. The method according to claim 1, 2, 3, or 8, wherein the complementary sequence of a coding sequence of said candidate sequence is administered to the host organism in an

amount sufficient to prevent the expression of a polypeptide encoded by said candidate sequence in said host organism.

37. The method according to claim 36, wherein said complementary sequence further comprises a cleaving moiety for cleaving RNA.

38. The method according to claim 1, 2, 3, or 8, wherein said candidate sequence is hybridized to nucleic acids from said host organism, and wherein the presence or absence of hybridization provides an indication of the presence or absence of said intracellular organism in a host cell from said host organism.

39. A system, comprising:

a) a first database comprising sequences from at least one host organism

b) a second database comprising genomic sequences from said host organism; and

c) an information management system comprising a search and subtraction function for eliminating sequences in said database comprising genomic sequences which are not found in said first database.

40. The system according to claim 39, further comprising at least one user device connectable to the network.

41. The system according to claim 39, wherein said system comprises a program capable of implementing an algorithm for comparing a plurality of sequences in the first database with all of the sequences in the second database.

42. The system according to claim 41, wherein said system comprises a MEGABLAST program.

43. The system according to claim 39, wherein said system comprises a high speed, linear array processor.

44. The system according to claim 39, wherein said system further comprises a result sequence set comprising sequences in the first database which do not match sequences in the genomic database.

45. The system according to claim 39, further comprising an identity matrix which requires a score of greater than or equal to 60.

46. The system according to claim 39 or 45, wherein the system iteratively computes the degree of alignment between sequences in the first and second database.

47. The system according to claim 45, wherein iterative computing is performed using progressively smaller word sizes.

48. The system according to claim 39, wherein the system provides one or more programs for performing one or more electronic subtraction functions for eliminating any of: vector sequences, repetitive sequences, mitochondrial sequences, sequences from non-host organisms, and combinations thereof, from the genomic database.

49. A computer program product comprising a computer readable memory on which is embedded one or more programs for implementing any of the system functions recited in claim 39 or 41.

50. A method of using a computer system to identify a microbe inhabiting a host organism, comprising the steps of:

obtaining sequence information from a plurality of expressed sequences from at least one host organism; and

searching a database of host organism genomic sequences to determine the presence or absence of the plurality of expressed sequences in the database, wherein the absence of an expressed sequence in the database identifies the expressed sequence as a candidate microbe sequence.

51. The method according to claim 50, wherein said plurality of sequences are from a library of sequences.

52. The method according to claim 51, wherein said library of sequences is a library of expressed sequences.

53. The method according to claim 51 or 52, wherein said library comprises human sequences.

54. The method according to claim 53, wherein said library comprises human sequences from one or more humans having a pathological condition.

55. The method according to claim 54, wherein said pathological condition is a disease selected from the group consisting of an inflammatory disease, an autoimmune disease, and a cell proliferative disease.

56. The method according to claim 55, wherein said disease is selected from the group consisting of sarcoidosis, inflammatory bowel disease, atherosclerosis, multiple sclerosis, rheumatoid arthritis, type I diabetes mellitus, lupus erythematosus, Hodgkin's disease, and bronchioalveolar carcinoma.

57. The method according to claim 50, wherein said step of obtaining sequence information comprises sequencing expressed sequences cloned in a library of expressed sequences.

58. A method of using a computer system to identify a microbe inhabiting a host organism, comprising the steps of:

obtaining expressed sequence information from a plurality of sequences from at least one non-microbial host organism; and

searching a database of microbial sequences to determine the presence or absence of the plurality of expressed sequences in the database, wherein the presence of an expressed sequence in the database identifies the expressed sequence as a candidate microbe sequence.

59. The method according to claim 58, wherein said plurality of sequences are from a library of expressed sequences.

60. The method according to claim 58, wherein said library of sequences comprises sequences from one or more humans having a pathological condition.

61. The method according to claim 60, wherein said pathological condition is an infectious disease.